This talk is based on the following paper:

Causal inference is a central goal of social science and public policy research.
Randomized experiments are seen as gold standard.
Design and analyze observational studies to replicate experiments.
But, experiments are a black box.
Can only tell whether the treatment causally affects the outcome.
Not how and why the treatment affects the outcome.
Qualitative research uses process tracing.

How can quantitative research be used to identify causal mechanisms?

What This Talk is About

Goal: Show how to identify causal mechanisms using statistics.
Method: Causal Mediation Analysis.

Direct and indirect effects; intermediate and intervening variables.
Popular among social psychologists (e.g., Baron and Kenny).
Regression

\[ Y_i = \alpha + \beta T_i + \gamma M_i + \delta X_i + \epsilon_i \]

- Each coefficient is interpreted as a causal effect
- Sometimes, it’s called marginal (or partial) effect
- Idea: increase \( T_i \) by one unit while holding \( M_i \) and \( X_i \) constant

Post-treatment bias: if you change \( T_i \), that may also change \( M_i \)

Usual advice: only include causally prior (or pre-treatment) variables

But, then you lose causal mechanisms!

Example I: Early Childhood Intervention Programs

- Do early childhood intervention programs have long-term impact on educational outcomes? If so, how?
- Observational data: The 1999 Chicago Longitudinal Study
- Low income predominantly African-American children in government-funded kindergarten programs

- Treatment: The Child-Parent Center preschool programs
  - structured half-day program for 3 and 4 years old
  - reading and writing activities
  - mandatory parental involvement

- Outcome: Participation in special education classes in grades 1–8
Hypotheses and Findings

- Mediators: cognitive advantage vs. family support

**Mediator: Cognitive Advantage**

- Treatment: Child-Parent Program
- Outcome: Special Education Placement

**Mediator: Parental Involvement**

**Findings:**
- The CPC program reduced the participation in special education
- Cognitive advantage mediates the effect of the program, but family support does not

---

Example II: Girls-Only Curriculum in Math

- Classroom environment and math achievement
- Do girl-only classrooms affect math learning? If so, how?
- Mediator: math anxiety

**Mediator: Math Anxiety**

- Treatment: All Girls Classroom
- Outcome: Math Performance/Interest
Design and Findings

- Two similar coeducational schools, one school offering girls-only math education
- Students filled out math attitudes surveys
- Outcomes: math achievement and sustained interest
- Large positive impact of girls-only curriculum
- Math anxiety does not seem to mediate this effect
- An alternative explanation: cooperative classroom environment?

### Formal Statistical Framework of Causal Inference

- **Units:** $i = 1, \ldots, n$
- **“Treatment”:** $T_i = 1$ if treated, $T_i = 0$ otherwise
- **Observed outcome:** $Y_i$
- **Pre-treatment covariates:** $X_i$
- **Potential outcomes:** $Y_i(1)$ and $Y_i(0)$ where $Y_i = Y_i(T_i)$

<table>
<thead>
<tr>
<th>Student</th>
<th>Teaching program</th>
<th>Post-test score</th>
<th>Gender</th>
<th>Pre-test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>$T_i$</td>
<td>$Y_i(1)$</td>
<td>$Y_i(0)$</td>
<td>$X_{1i}$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>88</td>
<td>?</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>?</td>
<td>76</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>?</td>
<td>85</td>
<td>F</td>
</tr>
<tr>
<td>...</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>$n$</td>
<td>1</td>
<td>89</td>
<td>?</td>
<td>M</td>
</tr>
</tbody>
</table>

- Causal effect: $Y_i(1) - Y_i(0)$
Identification of Causal Effects in Standard Settings

- **Nonparametric identification**: What can we learn from the data generating process alone?
- **Average Treatment Effect (ATE)**:
  \[ \tau \equiv \mathbb{E}(Y_i(1) - Y_i(0)) \]
- **Ignorability (randomization, no omitted variable)**:
  \[(Y_i(1), Y_i(0)) \perp \perp T_i \mid X_i \]
- **Identification under ignorability**:
  \[ \tau = \mathbb{E}(Y_i \mid T_i = 1, X_i) - \mathbb{E}(Y_i \mid T_i = 0, X_i) \]
- **Relationship with the linear regression**:
  \[ Y_i(T_i) = \alpha + \beta T_i + \gamma X_i + \epsilon_i \]
  where ignorability implies \( T_i \perp \perp \epsilon_i \mid X_i \)

Notation for Causal Mediation Analysis

- **Binary treatment**: \( T_i \in \{0, 1\} \)
- **Mediator**: \( M_i \)
- **Outcome**: \( Y_i \)
- **Observed covariates**: \( X_i \)
- **Potential mediators**: \( M_i(t) \) where \( M_i = M_i(T_i) \)
- **Potential outcomes**: \( Y_i(t, m) \) where \( Y_i = Y_i(T_i, M_i(T_i)) \)
Defining and Interpreting Causal Mediation Effects

- **Total causal effect**: \( \tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0)) \)

- **Causal mediation effects**:
  \[ \delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \]

  - Change the mediator from \( M_i(0) \) to \( M_i(1) \) while holding the treatment constant at \( t \)
  - Indirect effect of the treatment on the outcome through the mediator under treatment status \( t \)
  - \( Y_i(t, M_i(t)) \) is observable but \( Y_i(t, M_i(1 - t)) \) is not

- **Direct effects**:
  \[ \zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t)) \]

  - Change the treatment from 0 to 1 while holding the mediator constant at \( M_i(t) \)

  - Total effect = mediation (indirect) effect + direct effect:
    \[ \tau_i = \delta_i(t) + \zeta_i(1 - t) \]

- **Quantities of interest**: Average Causal Mediation Effects,
  \[ \bar{\delta}(t) \equiv \mathbb{E}(\delta_i(t)) = \mathbb{E}\{ Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \} \]
The Main Identification Result

Assumption 1 (Sequential Ignorability)

\[
\{ Y_i(t, m), M_i(t) \} \perp \perp T_i \mid X_i, \\
Y_i(t, m) \perp \perp M_i \mid T_i, X_i
\]

for \( t = 0, 1 \) and \( m \in \mathcal{M} \).

Theorem 1 (Nonparametric Identification)

Under Assumption 1, for \( t = 0, 1 \),

\[
\hat{\delta}(t) = (-1)^t \left\{ \int \mathbb{E}(Y_i \mid M_i, T_i = t, X_i) dP(M_i \mid T_i = 1 - t, X_i) \\
- \mathbb{E}(Y_i \mid T_i = t, X_i) \right\} dP(X_i)
\]

Theoretical and Practical Implications

- Existing statistics literature concludes that an additional assumption is required for the identification of mediation effects.
- However, sequential ignorability alone is sufficient.

- Fit two nonparametric regressions:
  1. \( \mu_{tm}(x) \equiv \mathbb{E}(Y_i \mid T_i = t, M_i = m, X_i = x) \)
  2. \( \lambda_{tm}(x) \equiv \text{Pr}(M_i = m \mid T_i = t, X_i = x) \)

- The plug-in estimator for a discrete mediator:

\[
\hat{\delta}(t) = \frac{(-1)^t}{n} \left\{ \sum_{i=1}^{n} \sum_{m=0}^{J-1} \hat{\mu}_{tm}(X_i) \left( \hat{\lambda}_{1-t,m}(X_i) - \hat{\lambda}_{tm}(X_i) \right) \right\}.
\]
Theorem 2 (Identification under LSEM)

Consider the following linear structural equation model

\[
M_i = \alpha_2 + \beta_2 T_i + \epsilon_{2i}, \\
Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i}.
\]

Under Assumption 1, the average causal mediation effects are identified as \( \bar{\delta}(0) = \bar{\delta}(1) = \beta_2 \gamma \).

- Run two (not three) regressions and multiply two coefficients!
- Direct effect: \( \beta_3 \)
- Total effect: \( \beta_2 \gamma + \beta_3 \)
- Total effect could be zero even when mediation effects are not

Need for Sensitivity Analysis

- The sequential ignorability assumption is often too strong
- Need to assess the robustness of findings via sensitivity analysis
- Question: How large a departure from the key assumption must occur for the conclusions to no longer hold?
- Parametric and nonparametric sensitivity analysis by assuming

\[
\{ Y_i(t, m), M_i(t) \} \perp T_i \mid X_i
\]

but not

\[
Y_i(t, m) \not\perp M_i \mid T_i, X_i
\]
Parametric Sensitivity Analysis

- **Sensitivity parameter**: \( \rho \equiv \text{Corr}(\epsilon_{2i}, \epsilon_{3i}) \)
- Existence of omitted variables leads to non-zero \( \rho \)
- Set \( \rho \) to different values and see how mediation effects change
- Can write estimated causal mediation effects as a function of \( \rho \) (and identifiable parameters)
- All you have to do: fit another regression

\[
Y_i = \alpha^*_3 + \beta^*_3 T_i + \epsilon^*_3 i
\]

in addition to the previous two regressions:

\[
M_i = \alpha_2 + \beta_2 T_i + \epsilon_{2i}
\]

\[
Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i}
\]

- When do my results go away completely?
- \( \bar{\delta}(t) = 0 \) if and only if \( \rho = \text{Corr}(\epsilon_{2i}, \epsilon^*_{3i}) \) (easy to compute!)

An Alternative Interpretation of \( \rho \)

- A common omitted variable \( U_i \)

\[
M_i = \alpha_2 + \beta_2 T_i + \lambda_2 U_i + \epsilon^*_{2i}
\]

\[
Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \lambda_3 U_i + \epsilon^*_{3i}
\]

- How much does \( U_i \) matter?

\[
R^2_M = 1 - \frac{\text{var}(\epsilon^*_{2i})}{\text{var}(\epsilon_{2i})} \quad \text{and} \quad R^2_Y = 1 - \frac{\text{var}(\epsilon^*_{3i})}{\text{var}(\epsilon_{3i})}.
\]

- The relationship:

\[
\rho^2 = R^2_M R^2_Y
\]
Concluding Remarks

- Quantitative analysis can be used to identify causal mechanisms!
- Estimate causal mediation effects rather than marginal effects
- Wide applications in social science disciplines

Contributions to the statistics literature:

1. Clarify assumptions
2. Extend parametric method
3. Develop nonparametric method
4. Provide new sensitivity analysis